

Robust Consensus in the Presence of Impulsive Channel Noise

Sivaraman Dasarathan, Cihan Tepedelenlioğlu, *Member, IEEE*, Mahesh Banavar, *Member, IEEE* and Andreas Spanias, *Fellow, IEEE*

Abstract

A distributed average consensus algorithm robust to a wide range of impulsive channel noise distributions is proposed. This work is the first of its kind in the literature to propose a consensus algorithm which relaxes the requirement of finite moments on the communication noise. It is shown that the nodes reach consensus asymptotically to a finite random variable whose expectation is the desired sample average of the initial observations with a variance that depends on the step size of the algorithm and the receiver nonlinear function. The asymptotic performance is characterized by deriving the asymptotic covariance matrix using results from stochastic approximation theory. Simulations corroborate our analytical findings and highlight the robustness of the proposed algorithm.

Index Terms

Distributed Consensus, Sensor Networks, Bounded Transmissions, Impulsive Noise, Asymptotic Covariance, Stochastic Approximation, Markov Processes.

I. INTRODUCTION

Wireless sensor networks (WSNs) without a fusion center have the advantages of robustness to node failures and being able to function autonomously without a central node controlling the entire network [1]. In such fully distributed networks, sensors collaborate with their neighbours by repeatedly exchanging information which they combine locally to achieve a desired global

The authors are with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA. (Email: {sdasarat, cihan, mbanavar, spanias}@asu.edu). This work was supported in part by the National Science Foundation under Grants NSF FRP 1231034 and NSF CCSS 1307982.

objective. For example, the sensors could come to an agreement on the sample average (or on a global function) of initial measurements. This is called distributed consensus. Distributed consensus algorithms have attracted significant interest in the recent past and have found several applications in areas such as healthcare, environmental monitoring, military and home appliances [2]–[8].

In existing literature on consensus in the presence of communication noise, the additive noise is always assumed to have finite moments [6], [8]–[16]. Sensor networks which operate in adverse conditions can be susceptible to impulsive noise distributions. For example, the aggregated interference at a desired node from its neighbouring nodes of a Poisson network is characterized by alpha-stable distribution which may not have finite mean or variance [17]–[24]. Therefore there is a need to develop consensus algorithms which are robust to impulsive channel noise. Consensus with nonlinear combining at the receiver has been considered in [4], [25]–[29] only in the absence of inter-sensor communication noise. Therefore, it is of interest to solve the problem of distributed consensus with receiver nonlinearities that soft-limit the impulsive additive noise.

In this paper, we propose a robust consensus (RC) algorithm which is robust to impulsive communication noise by soft-limiting at receiver sensor nodes before combining. We do not require the channel noise to have finite moments as is assumed in all the previous work on distributed average consensus algorithms [6], [8]–[16]. In addition, like in [30], we assume that every sensor maps its state value through a bounded function before transmission to respect a peak power constraint at every iteration making it ideal for resource-constrained WSNs. We prove that all the sensors employing the RC algorithm reach consensus to a finite random variable whose mean is the desired sample average. We characterize the asymptotic performance by deriving the asymptotic covariance matrix using results from stochastic approximation theory. Finally, we explore the performance of the proposed algorithm employing various functions for the transmit and receiver non-linearities. Different from [6], [8] and [16] which also considered consensus in the presence of noisy transmissions, herein we analyse nonlinear processing both at the transmit and receiver nodes and study the asymptotic covariance matrix and its dependence on both the power-constraining transmit nonlinearity, and the soft-limiting receive nonlinearity. It is shown that the norm of the asymptotic covariance matrix is limited by the Fisher information of the noise distribution with respect to a location parameter.

The rest of this paper is organized as follows. We begin by reviewing network graph theory

in Section II. In Section III, we describe the sensing and channel models and introduce the consensus problem. We consider the RC algorithm in the presence of noise in Section IV, and prove that the sensors reach consensus to a random variable. In Section V, we present several simulation examples to study the performance of the proposed algorithm. Concluding remarks are presented in Section VI.

Notations and Conventions

Vectors are denoted by boldface upper-case or lower-case letters and matrices are denoted by boldface upper-case letters. $\max\{a_1, a_2\}$ denotes the maximum of a_1 and a_2 . $\text{diag}[a_1, a_2, \dots, a_N]$ denotes an $N \times N$ diagonal matrix whose diagonal elements are given by a_1, a_2, \dots, a_N . $\mathbb{E}[\cdot]$ denotes the expectation operator. The symbol $\|\cdot\|$ denotes the l_2 norm for vectors and spectral norm for symmetric matrices. For a symmetric matrix \mathbf{M} , $\lambda_i(\mathbf{M})$, $i = 1, \dots, N$, denotes the i^{th} smallest eigenvalue, $\mathbf{1} := [1 \ 1 \dots 1]^T$, and \mathbf{I} denotes the identity matrix.

II. REVIEW OF NETWORK GRAPH THEORY

In this section, we provide a brief background on network graph theory. Consider an undirected graph $\mathbb{G} = (\mathbb{N}, \mathbb{E})$ containing a set of nodes $\mathbb{N} = \{1, \dots, N\}$ and a set of edges \mathbb{E} . Nodes that communicate with each other have an edge between them. We denote the set of neighbours of node i by \mathbb{N}_i , $\mathbb{N}_i = \{j | \{i, j\} \in \mathbb{E}\}$ where $\{i, j\}$ indicates an edge between the nodes i and j [31]. A graph is connected if there exists at least one path between every pair of nodes. We denote the number of neighbours of a node i by d_i and $d_{\max} = \max_i d_i$. The graph structure is described by an $N \times N$ symmetric matrix called the adjacency matrix \mathbf{A} , whose i, j element $[\mathbf{A}]_{i,j} = 1$ if $\{i, j\} \in \mathbb{E}$. The diagonal matrix $\mathbf{D} = \text{diag}[d_1, d_2, \dots, d_N]$ captures the degrees of all the nodes in the network. The Laplacian matrix of the graph is defined as $\mathbf{L} := \mathbf{D} - \mathbf{A}$. The graph Laplacian characterises a number of useful properties of the graph. The eigenvalues of \mathbf{L} are non-negative and the number of zero eigenvalues denotes the number of distinct components of the graph. When the graph is connected, $\lambda_1(\mathbf{L}) = 0$, and $\lambda_i(\mathbf{L}) > 0, i \geq 2$, so that the rank of \mathbf{L} for a connected graph is $N - 1$. The vector $\mathbf{1}$ is the eigenvector of \mathbf{L} associated with the eigenvalue 0, i.e., $\mathbf{L}\mathbf{1} = \mathbf{0}$. The eigenvalue $\lambda_2(\mathbf{L})$ characterizes how densely the graph is connected and the performance of consensus algorithms depend on this eigenvalue [32].

III. SENSING AND CHANNEL MODEL

A. Sensing Model

Consider a WSN with N sensor nodes each with an initial measurement $x_i(0) \in \mathbb{R}$, $i = 1, \dots, N$. Measurements made at the sensor nodes are modeled as

$$x_i(0) = \theta + \eta_i, \quad i = 1, \dots, N \quad (1)$$

where θ is an unknown real-valued parameter and η_i is the sensing noise at the i^{th} sensor. For many distributions on η_i , the sample mean of these initial measurements is the maximum likelihood estimate of θ :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i(0). \quad (2)$$

We would like to design an iterative distributed algorithm, in which each sensor communicates only with its neighbours and each sensor has a state that converges to \bar{x} . If the states of all the sensor nodes converge to \bar{x} , then the network is said to have reached *consensus* on the sample average.

B. Channel Model

Each sensor can transmit or receive information to or from its neighbours. When a sensor transmits its state information, it can send a function of its state instead of the state itself. In this link there is additive noise at the receiver node which can be modeled as

$$y_{ij}(t) = h(x_j(t)) + n_{ij}(t), \quad \{i, j\} \in \mathbb{E}, \quad (3)$$

where $x_j(t)$, $j \in \mathbb{N}_i$, is the state value of the j^{th} node at time t ; $h(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is the power-constraining transmission function used at every node, $n_{ij}(t)$ is the noise associated with the reception of $h(x_j(t))$, and $y_{ij}(t)$ is the received signal at node i from node j at time t . The existing linear consensus algorithms in [6], [8]–[16] require $n_{ij}(t)$ to have finite moments. Instead, we assume that the noise samples $n_{ij}(t)$ are mutually independent identically distributed (i.i.d.), symmetric real-valued with zero median (e.g., its PDF, when it exists, is symmetric about zero).

IV. ROBUST CONSENSUS WITH IMPULSIVE COMMUNICATION NOISE

In this section, we propose a robust consensus algorithm in which every node performs a nonlinear operation by soft-limiting the noisy state information at the receiver node. The receiver non-linearity makes the algorithm robust to a wide range of heavy-tailed channel noise distributions. Also, at the transmitter side every sensor maps its state value through a bounded function before transmission to constrain the transmit power making it ideal for resource-constrained WSNs.

A. The RC Algorithm with Communication Noise

As discussed in (3), each sensor maps its state value at time t through the function $h(x)$ before transmission, and combines the received state values through a nonlinear function $f(x)$ according to the following recursion:

$$\begin{aligned} x_i(t+1) &= x_i(t) - \alpha(t) \sum_{j \in \mathbb{N}_i} [f(h(x_i(t))) - y_{ij}(t)] , \\ &= x_i(t) - \alpha(t) \sum_{j \in \mathbb{N}_i} [f(h(x_i(t))) - h(x_j(t)) - n_{ij}(t)] , \end{aligned} \quad (4)$$

where $i = 1, \dots, N$, and $t = 0, 1, 2, \dots$, is the time index, and $\alpha(t)$ is a positive step size which will be assumed to satisfy assumption **(A5)** in the sequel. The node j transmits its information $x_j(t)$ by mapping it through the function $h(x)$, node i receives a noisy signal $h(x_j(t)) + n_{ij}(t)$. The function $f(x)$ is applied at the receiver side to combat the effect of impulsive channel noise $n_{ij}(t)$ and will be further assumed to satisfy **(A2)** in the sequel.

We now compare the existing work on nonlinear consensus in [4], [25]–[29] against the proposed algorithm in (4). The algorithm in [25] becomes a special case of (4) with $h(x) = x$ and $f(x) = \sin(x)$ in a setting with no channel noise ($n_{ij}(t) \equiv 0$). The algorithm in [4] becomes a special case of (4) with $h(x) = x$ and $f(x)$ being an increasing odd function. There is no communication noise assumed in all the existing work on consensus with nonlinear $f(\cdot)$ [4], [25]–[29] whereas we consider herein the communication noise in the presence of both the transmit and receive non-linearities. Moreover, with the transmit non-linearity $h(x)$, the transmit power from all the sensors are always bounded which is a desirable feature for power constrained WSNs. The NLC algorithm considered in [30] is a special case of (4) with $f(x) = x$ but assumes noise samples have finite moments, and fails in the presence of impulsive channel noise.

We make the following assumptions on $f(x)$, $h(x)$, $n_{ij}(t)$, $\alpha(t)$ and the graph:

Assumptions

(A1) Graph: The graph \mathbb{G} is undirected and connected so that $\lambda_2(\mathbf{L}) > 0$ [31].

(A2) Receive Nonlinearity: The function $f(x)$ is strictly increasing, odd and bounded.

(A3) Transmit Nonlinearity: The function $h(x)$ is strictly increasing.

(A4) Independent Noise Sequence: The noise samples $n_{ij}(t)$ are mutually i.i.d., symmetric real-valued with zero median (e.g., its PDF, when it exists, is symmetric about zero).

(A5) Decreasing Weight Sequence: In order to control the variance growth rate of the cumulative noise we need the following standard conditions on the sequence $\alpha(t)$:

$$\alpha(t) > 0, \sum_{t=0}^{\infty} \alpha(t) = \infty, \sum_{t=0}^{\infty} \alpha^2(t) < \infty. \quad (5)$$

Let $g(x) : \mathbb{R} \rightarrow \mathbb{R}$ be such that $g(x) := \mathbb{E}_n[f(x+n)]$ where $\mathbb{E}_n[\cdot]$ denotes the expectation with respect to any of the i.i.d. $n_{ij}(t)$ so that $f(x+n) = g(x) + v(x,n)$. Here $v(x,n) = f(x+n) - \mathbb{E}_n[f(x+n)]$ is a noise process which depends on $x \in \mathbb{R}$ and its randomness is due to the noise process n , and satisfies $\mathbb{E}_n[v(x,n)] = 0, x \in \mathbb{R}$. Let $\sigma^2 := \sup_x \text{var}[f(x+n)]$. Since $f(\cdot)$ is bounded due to **(A2)**, σ^2 is finite. Hence we have $\text{var}[f(x+n)] = \text{var}[v(x,n)] = \mathbb{E}[v^2(x,n)] \leq \sigma^2$. Using the fact that $f(x)$ is a strictly increasing odd function and that $-n$ has the same distribution as n due to symmetry, it can be easily proved that $g(x)$ is a strictly increasing odd function satisfying $g(0) = 0$. Using $g(x)$, the recursion in (4) can be written as

$$x_i(t+1) = x_i(t) - \alpha(t) \sum_{j \in \mathbb{N}_i} [g(h(x_i(t)) - h(x_j(t))) + v(h(x_i(t)) - h(x_j(t)), n_{ij}(t))] . \quad (6)$$

The recursion in (6) can be written in vector form as

$$\mathbf{X}(t+1) = \mathbf{X}(t) - \alpha(t) [\boldsymbol{\mu}(\mathbf{X}(t)) + \mathbf{n}(t, \mathbf{X}(t))] , \quad (7)$$

where $\mathbf{X}(t) \in \mathbb{R}^N$ is the state vector at time t given by $\mathbf{X}(t) = [x_1(t) \ x_2(t) \ \dots \ x_N(t)]^T$, and $\boldsymbol{\mu}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a function with i^{th} element is given by

$$[\boldsymbol{\mu}(\mathbf{x})]_i = \sum_{j \in \mathbb{N}_i} g(h(x_i) - h(x_j)) , 1 \leq i \leq N , \quad (8)$$

and $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_N]^T$. Due to the fact that $g(x)$ is odd and that the graph is connected, we have $\mathbf{1}^T \boldsymbol{\mu}(\mathbf{x}) = 0$. The vector $\mathbf{n}(t, \mathbf{X}(t))$ in (7) captures the additive noise at N nodes contributed

by their neighbours and their state values and its i^{th} component is given by

$$[\mathbf{n}(t, \mathbf{X}(t))]_i = - \sum_{j \in \mathbb{N}_i} v(h(x_i(t)) - h(x_j(t)), n_{ij}(t)) , 1 \leq i \leq N . \quad (9)$$

Clearly, conditioned on $\mathbf{X}(t) = \mathbf{x}$, the noise $\{v(h(x_i) - h(x_j), n_{ij}(t))\}_{t \geq 0, 1 \leq i, j \leq N}$ is an independent sequence across time t , and sensors i due to assumption **(A4)**. It also satisfies

$$\mathbb{E}[\mathbf{n}(t, \mathbf{x})] = \mathbf{0} , \forall t, \mathbf{x} , \quad \varrho := \sup_{t, \mathbf{x}} \mathbb{E}[\|\mathbf{n}(t, \mathbf{x})\|^2] \leq Nd_{\max}\sigma^2 < \infty. \quad (10)$$

Note that the inequality in (10) is because of **(A2)** and the fact that the number of neighbours of a given node is upper bounded by d_{\max} .

We will prove convergence of the RC algorithm in Section IV-B and asymptotic normality in Section IV-D. We now present a result on the convergence of a discrete time Markov process which will be used in establishing convergence of the RC algorithm.

B. A Result on the Convergence of Discrete time Markov Processes

Let $\mathcal{X} = \{\mathbf{X}(t)\}_{t \geq 0}$ be a discrete time vector Markov process on \mathbb{R}^N . The generating operator \mathcal{L} of \mathcal{X} is defined as

$$\mathcal{L}V(\mathbf{x}) = \mathbb{E}[V(\mathbf{X}(t+1)) | \mathbf{X}(t) = \mathbf{x}] - V(\mathbf{x}) \quad (11)$$

for functions $V(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$, for which the conditional expectation exists. Let $\mathbb{B} \subset \mathbb{R}^N$ and its complement be $\mathbb{B}' = \mathbb{R}^N \setminus \mathbb{B}$. We now state the desired result as a simplification of Theorem 2.7.1 in [33] (see also Theorem 1 in [8]). In general $\mathcal{L}V(\mathbf{x})$ may depend on t .

Theorem 1. *Let \mathcal{X} be a discrete time vector Markov process with the generator operator \mathcal{L} as in (11). If there exists a potential function $V(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^+$, and $\mathbb{B} \subset \mathbb{R}^N$ with the following properties*

$$V(\mathbf{x}) > 0, \mathbf{x} \in \mathbb{B}', \quad V(\mathbf{x}) = 0, \mathbf{x} \in \mathbb{B} , \quad (12)$$

$$\mathcal{L}V(\mathbf{x}) \leq -\gamma(t)\varphi(\mathbf{x}) + m\zeta(t)[1 + V(\mathbf{x})] \quad (13)$$

where $m > 0$, $\varphi(\mathbf{x})$ is such that

$$\varphi(\mathbf{x}) = 0, \mathbf{x} \in \mathbb{B}, \quad \varphi(\mathbf{x}) > 0, \mathbf{x} \in \mathbb{B}' , \quad (14)$$

and

$$\gamma(t) > 0, \zeta(t) > 0, \sum_{t=0}^{\infty} \gamma(t) = \infty, \sum_{t=0}^{\infty} \zeta(t) < \infty, \quad (15)$$

then, the discrete time vector Markov process $\mathcal{X} = \{\mathbf{X}(t)\}_{t \geq 0}$ with arbitrary initial distribution converges almost surely (a.s.) to the set \mathbb{B} as $t \rightarrow \infty$. That is,

$$\Pr \left[\liminf_{t \rightarrow \infty} \inf_{\mathbf{z} \in \mathbb{B}} \|\mathbf{X}(t) - \mathbf{z}\| = 0 \right] = 1. \quad (16)$$

Intuitively, Theorem 1 indicates that if the one-step prediction error of the Markov process evaluated at the potential function in (11) is bounded as in (13) then it is possible to establish convergence of $\mathbf{X}(t)$.

To prove the a.s. convergence of the consensus algorithm in (7) using Theorem 1, we choose the consensus subspace \mathbb{B} , the set of all vectors whose entries are of equal value as,

$$\mathbb{B} = \{\mathbf{x} \in \mathbb{R}^N | \mathbf{x} = a\mathbf{1}, a \in \mathbb{R}\}. \quad (17)$$

We are now ready to state the main result of Section IV. But first, we start out with a preparatory lemma.

Lemma 1. Define a positive semi-definite matrix \mathbf{M} as the Laplacian of a fully connected graph: $\mathbf{M} := N\mathbf{I} - \mathbf{1}\mathbf{1}^T$. Let $\mathbf{x} \in \mathbb{B}'$, then $\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) > 0$.

Proof: Consider

$$\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) = \mathbf{x}^T [N\mathbf{I} - \mathbf{1}\mathbf{1}^T] \boldsymbol{\mu}(\mathbf{x}), \quad (18)$$

$$= N\mathbf{x}^T \boldsymbol{\mu}(\mathbf{x}) - \mathbf{x}^T \mathbf{1}\mathbf{1}^T \boldsymbol{\mu}(\mathbf{x}), \quad (19)$$

$$= N\mathbf{x}^T \boldsymbol{\mu}(\mathbf{x}), \quad (20)$$

where we have used the fact that $\mathbf{1}^T \boldsymbol{\mu}(\mathbf{x}) = 0$ in (19) to get (20). Expanding $\mathbf{x}^T \boldsymbol{\mu}(\mathbf{x})$ using (8), we get

$$\begin{aligned} \mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) = N \left[\sum_{j \in \mathbb{N}_1} g(h(x_1) - h(x_j))x_1 + \sum_{j \in \mathbb{N}_2} g(h(x_2) - h(x_j))x_2 \right. \\ \left. + \dots + \sum_{j \in \mathbb{N}_N} g(h(x_N) - h(x_j))x_N \right]. \end{aligned} \quad (21)$$

Note that the i^{th} summation in (21) corresponds to the i^{th} node. Now suppose that node i is connected to node j . Then there exists a term $g(h(x_i) - h(x_j))x_i$ in the summation corresponding to the i^{th} node in (21), and a term $g(h(x_j) - h(x_i))x_j$ in the summation corresponding to the j^{th} node in (21). Both of these terms can be combined as $(x_i - x_j)g(h(x_i) - h(x_j))$ and this corresponds to the edge $\{i, j\} \in \mathbb{E}$. Thus equation (21) can be written as pairwise products enumerated over all the edges in the graph as follows

$$\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) = N \sum_{\{i,j\} \in \mathbb{E}} (x_i - x_j) g(h(x_i) - h(x_j)) . \quad (22)$$

Since $\mathbf{x} \in \mathbb{B}'$, $\varphi(\mathbf{x})$ in (22) is positive due to the facts that $h(x)$ is strictly increasing and $g(x)$ is a strictly increasing odd function so that there is at least one term in the sum which is greater than zero and this completes the proof. \blacksquare

Theorem 2. *Let the assumptions (A1)-(A5) hold. Consider the RC algorithm in (7) with the initial state vector $\mathbf{X}(0) \in \mathbb{R}^N$. Then, the state vector $\mathbf{X}(t)$ in (7) approaches the consensus subspace \mathbb{B} a.s., i.e.,*

$$\Pr \left[\liminf_{t \rightarrow \infty} \inf_{\mathbf{z} \in \mathbb{B}} \|\mathbf{X}(t) - \mathbf{z}\| = 0 \right] = 1. \quad (23)$$

Proof: We will make use of Theorem 1 to prove (23). We will choose an appropriate potential function $V(\mathbf{x})$ that is non-negative and satisfies equation (12). We will then prove that the generating operator \mathcal{L} applied on $V(\mathbf{x})$ as in (11) can be upper bounded as in (13) with $\gamma(t) = \alpha(t)$, $\zeta(t) = \alpha^2(t)$, and a $\varphi(\mathbf{x})$ will be chosen to satisfy (14).

First we see that under the assumptions the discrete time vector process $\{\mathbf{X}(t)\}_{t \geq 0}$ in (7) is Markov. Let \mathbf{M} be a positive semi-definite matrix as defined in Lemma 1. Let $V(\mathbf{x}) = \mathbf{x}^T \mathbf{M} \mathbf{x}$, then the function $V(\mathbf{x})$ is non-negative since \mathbf{M} is a positive semi-definite matrix. Note that any $\mathbf{x} \in \mathbb{B}$ is an eigenvector of \mathbf{M} associated with the zero eigenvalue, therefore we have

$$V(\mathbf{x}) = 0, \mathbf{x} \in \mathbb{B} . \quad (24)$$

We have now verified that $V(\mathbf{x})$ satisfies the second condition in (12). We now proceed to show the first condition. Let $\mathbf{x} = \mathbf{x}_{\mathbb{B}} + \mathbf{x}_{\mathbb{B}^\perp}$ where $\mathbf{x}_{\mathbb{B}}$ is the orthogonal projection of \mathbf{x} on \mathbb{B} . When $\mathbf{x} \in \mathbb{B}'$, we have $\|\mathbf{x}_{\mathbb{B}^\perp}\| > 0$. Therefore, for any $\mathbf{x} \in \mathbb{B}'$,

$$V(\mathbf{x}) = V(\mathbf{x}_{\mathbb{B}}) + V(\mathbf{x}_{\mathbb{B}^\perp}) = V(\mathbf{x}_{\mathbb{B}^\perp}) \geq \lambda_2(\mathbf{M}) \|\mathbf{x}_{\mathbb{B}^\perp}\|^2 > 0 , \quad (25)$$

where the last inequality is due to $\lambda_2(\mathbf{M}) > 0$. The equations (24) and (25) establish that the conditions in (12) in Theorem 1 are satisfied.

Let $\mathbf{x} \in \mathbb{B}'$ and $\boldsymbol{\mu}(\mathbf{x})$ be as defined in (8), and $\boldsymbol{\mu}_{\mathbb{B}}(\mathbf{x})$ be the orthogonal projection of $\boldsymbol{\mu}(\mathbf{x})$ on \mathbb{B} . Then, $\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\mu}_{\mathbb{B}}(\mathbf{x}) + \boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})$, where $\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})$ is non-zero, i.e., $\|\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})\| > 0$ which is proved now. First we recall that $\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) > 0$ when $\mathbf{x} \in \mathbb{B}'$ due to Lemma 1. This means $(\mathbf{x}_{\mathbb{B}} + \mathbf{x}_{\mathbb{B}\perp})^T \mathbf{M}(\boldsymbol{\mu}_{\mathbb{B}}(\mathbf{x}) + \boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})) = \mathbf{x}_{\mathbb{B}\perp}^T \mathbf{M} \boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x}) > 0$ for $\mathbf{x} \in \mathbb{B}'$. If $\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})$ were zero, then $\mathbf{x}_{\mathbb{B}\perp}^T \mathbf{M} \boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x}) = 0$ which contradicts with the fact that $\mathbf{x}_{\mathbb{B}\perp}^T \mathbf{M} \boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x}) > 0$. Therefore, $\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})$ is non-zero. Define $\beta := \sup_{\mathbf{x}} \|\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})\|^2 / \|\mathbf{x}_{\mathbb{B}\perp}\|^2$, then $0 < \beta < \infty$, where the finiteness of β can be seen from the fact that $\boldsymbol{\mu}(\mathbf{x})$ is bounded for all \mathbf{x} because $f(x)$ is bounded due to (A2), and by expressing $\boldsymbol{\mu}(\mathbf{x})$ around $\mathbf{x} = a\mathbf{1}$, $a \in \mathbb{R}$ using Taylor's series and observing that the ratio $\|\boldsymbol{\mu}(\mathbf{x})\|^2 / \|\mathbf{x}_{\mathbb{B}\perp}\|^2$ is finite as $\mathbf{x} \rightarrow a\mathbf{1}$.

Now we will prove that (13) is satisfied as well. Towards this end, consider $\mathcal{L}V(\mathbf{x})$ defined in (11),

$$\mathcal{L}V(\mathbf{x}) = \mathbb{E} \left[\mathbf{X}(t+1)^T \mathbf{M} \mathbf{X}(t+1) | \mathbf{X}(t) = \mathbf{x} \right] - V(\mathbf{x}) , \quad (26)$$

$$\begin{aligned} &= \mathbb{E} \left[(\mathbf{x}^T - \alpha(t) (\boldsymbol{\mu}(\mathbf{x})^T + \mathbf{n}^T(t, \mathbf{x}))) \cdot (\mathbf{M} \mathbf{x} - \alpha(t) (\mathbf{M} \boldsymbol{\mu}(\mathbf{x}) + \mathbf{M} \mathbf{n}(t, \mathbf{x}))) \right] \\ &\quad - V(\mathbf{x}) , \end{aligned} \quad (27)$$

$$= -2\alpha(t) \mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) + \alpha^2(t) \boldsymbol{\mu}(\mathbf{x})^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) + \mathbb{E} [\mathbf{n}^T(t, \mathbf{x}) \mathbf{M} \mathbf{n}(t, \mathbf{x})] . \quad (28)$$

We get (28) by expanding (27) and taking the expectations and using the fact that $\mathbb{E}[\mathbf{n}(t, \mathbf{x})] = \mathbf{0}$.

We have

$$\mathbb{E} [\mathbf{n}^T(t, \mathbf{x}) \mathbf{M} \mathbf{n}(t, \mathbf{x})] \leq \mathbb{E} [\lambda_N(\mathbf{M}) \|\mathbf{n}(t, \mathbf{x})\|^2] \leq \lambda_N(\mathbf{M}) \varrho , \quad (29)$$

where the second inequality follows from (10). Using (29) in (28), we get the following bound

$$\mathcal{L}V(\mathbf{x}) \leq -2\alpha(t) [\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x})] + \alpha^2(t) [\boldsymbol{\mu}(\mathbf{x})^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) + \varrho \lambda_N(\mathbf{M})] , \quad (30)$$

$$\leq -2\alpha(t) [\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x})] + \alpha^2(t) [\lambda_N(\mathbf{M}) \beta \|\mathbf{x}_{\mathbb{B}\perp}\|^2 + \varrho \lambda_N(\mathbf{M})] , \quad (31)$$

$$\leq -2\alpha(t) [\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x})] + \alpha^2(t) [\beta \mathbf{x}^T \mathbf{M} \mathbf{x} + \varrho N] , \quad (32)$$

$$\leq -2\alpha(t) [\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x})] + m\alpha^2(t) [1 + \beta_2 \mathbf{x}^T \mathbf{M} \mathbf{x}] , \quad (33)$$

$$\leq -\alpha(t) \varphi(\mathbf{x}) + m\alpha^2(t) [1 + V(\mathbf{x})] , \quad (34)$$

where we have used the fact $\boldsymbol{\mu}(\mathbf{x})^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) \leq \lambda_N(\mathbf{M}) \|\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})\|^2$ and $\|\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})\|^2 \leq \beta \|\mathbf{x}_{\mathbb{B}\perp}\|^2$ in (30) to get (31). In (31), we have used the fact that $\mathbf{x}^T \mathbf{M} \mathbf{x} \geq \lambda_2(\mathbf{M}) \|\mathbf{x}_{\mathbb{B}\perp}\|^2$ due to (25) and

$\lambda_2(\mathbf{M}) = \lambda_N(\mathbf{M}) = N$ to get (32). In (32), we defined $m := \max\{\beta, \varrho N\}$, and $\beta_2 := \varrho N/m$ to get (33) and it is easy to see that $\beta_2 \in (0, 1]$. From (33), due to the fact that $\beta_2 \in (0, 1]$ and letting $\varphi(\mathbf{x}) := 2\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x})$, we get (34).

We will now prove that $\varphi(\mathbf{x})$ in (34) satisfies equation (14) of Theorem 1.

Whenever $\mathbf{x} \in \mathbb{B}$, i.e., $\mathbf{x} = a\mathbf{1}$, $a \in \mathbb{R}$, then $x_i = x_j, \forall i, j$, which means $g(h(x_i) - h(x_j)) = g(0) = 0, \forall i, j$, and hence $\boldsymbol{\mu}(\mathbf{x}) = \mathbf{0}$. This implies that $\varphi(\mathbf{x}) = 0, \forall \mathbf{x} \in \mathbb{B}$. From Lemma 1, it is immediate that $\varphi(\mathbf{x}) = 2\mathbf{x}^T \mathbf{M} \boldsymbol{\mu}(\mathbf{x}) > 0$ whenever $\mathbf{x} \in \mathbb{B}'$.

Letting $\gamma(t) = \alpha(t), \zeta(t) = \alpha^2(t)$ and by assumption **(A5)**, we see that the sequence $\alpha(t)$ in (34) satisfies (15). Thus all the conditions of Theorem 1 are satisfied to yield (23). ■

Theorem 2 states that the sample paths of $\mathbf{X}(t)$ approach the consensus subspace almost surely. Now, like in [8], we will prove the convergence of $\mathbf{X}(t)$ to a finite point in \mathbb{B} in Theorem 3.

Theorem 3. *Let the assumptions of Theorem 2 hold. Consider the RC algorithm in (7) with the initial state $\mathbf{X}(0) \in \mathbb{R}^N$. Then, there exists a finite real random variable θ^* such that*

$$\Pr \left[\lim_{t \rightarrow \infty} \mathbf{X}(t) = \theta^* \mathbf{1} \right] = 1. \quad (35)$$

Proof: Let the average of $\mathbf{X}(t)$ be $\bar{x}(t) = \mathbf{1}^T \mathbf{X}(t)/N$. It suffices to show that $\{\bar{x}(t)\}_{t \geq 0}$ is an \mathcal{L}_2 bounded martingale. A sequence of random variables $\{y(t)\}_{t \geq 0}$ is called as a martingale if for all $t \geq 0$, $E[|y(t)|] < \infty$ and $E[y(t+1) | y(1) y(2) \dots y(t)] = y(t)$. The sequence $\{y(t)\}_{t \geq 0}$ is an \mathcal{L}_2 bounded martingale if $\sup_t E[y^2(t)] < \infty$ (see [34, pp. 110]). Since $\mathbf{1}\bar{x}(t) \in \mathbb{B}$, Theorem 2 implies,

$$\Pr \left[\lim_{t \rightarrow \infty} \|\mathbf{X}(t) - \bar{x}(t)\mathbf{1}\| = 0 \right] = 1, \quad (36)$$

where (36) follows from (23) since the infimum in (23) is achieved by $\mathbf{z} = \bar{x}(t)\mathbf{1}$. Pre-multiplying (7) by $\mathbf{1}^T/N$ on both sides and noting that $\mathbf{1}^T \boldsymbol{\mu}(\mathbf{x}) = 0, \forall \mathbf{x}$ due to the symmetric structure of the graph we get,

$$\bar{x}(t+1) = \bar{x}(t) - \tilde{v}(t) \quad (37)$$

$$= \bar{x}(0) - \sum_{0 \leq k \leq t} \tilde{v}(k) \quad (38)$$

where $\tilde{v}(t) = \alpha(t) \mathbf{1}^T \mathbf{n}(t, \mathbf{X}(t)) / N$. From (10) it follows that

$$\begin{aligned} \mathbb{E}[\tilde{v}(t)] &= 0, \\ \sum_{t \geq 0} \mathbb{E}[\tilde{v}^2(t)] &= \sum_{t \geq 0} \frac{\alpha^2(t)}{N^2} \mathbb{E}[\|\mathbf{n}(t, \mathbf{X}(t))\|^2] \leq \frac{\varrho}{N^2} \sum_{t \geq 0} \alpha^2(t) < \infty \end{aligned}$$

which implies

$$\mathbb{E}[\bar{x}^2(t+1)] \leq \bar{x}^2(0) + \frac{\varrho}{N^2} \sum_{t \geq 0} \alpha^2(t), \forall t. \quad (39)$$

Equation (39) together with (37) implies that the sequence $\{\bar{x}(t)\}_{t \geq 0}$ is an \mathcal{L}_2 bounded martingale and hence converges a.s. to a finite random variable θ^* (see [33, Theorem 2.6.1]). Therefore the theorem follows from (36). \blacksquare

In what follows, we present the properties of the limiting random variable θ^* .

C. Mean Square Error of RC Algorithm

Theorems 2 and 3 establish that the sensors reach consensus asymptotically and converge a.s. to a finite random variable θ^* . We can view θ^* as an estimate of \bar{x} . In the following theorem we characterize the bias and mean squared error (MSE) properties of θ^* . We define the MSE of θ^* as $\xi_N = \mathbb{E}[(\theta^* - \bar{x})^2]$.

Theorem 4. *Let θ^* be the limiting random variable as in Theorem 3. Then θ^* is unbiased, $\mathbb{E}[\theta^*] = \bar{x}$, and its MSE is bounded, $\xi_N \leq \varrho N^{-2} \sum_{t \geq 0} \alpha^2(t)$.*

The proof is obtained by following the same steps of the Lemma 5 in [8].

We point out that with non-linear processing at both the transmitter and receiver nodes, we have obtained a similar bound on the MSE ξ_N as that of the linear consensus algorithm in [8] but in our case the bound depends on the function $f(x)$ (see assumption **(A2)**) through ϱ but does not depend on $h(x)$. Recall that $\varrho \leq N d_{\max} \sigma^2$ from (10) which implies that $\xi_N \leq d_{\max} N^{-1} \sum_{t \geq 0} \alpha^2(t) \sigma^2$. Therefore, if d_{\max} is finite for a large connected network, we have $\lim_{N \rightarrow \infty} \xi_N = 0$ and this means that θ^* converges to \bar{x} for large N . If the graph is densely connected, then d_{\max} is relatively high which increases the worst-case MSE. On the other hand, when the graph is densely connected, $\lambda_2(\mathbf{L})$ is larger which aids in the speed of convergence to θ^* , as quantified through the covariance matrix in Section IV-D.

D. Asymptotic Normality of RC Algorithm

In this section, we establish the asymptotic normality of the RC algorithm in (7). Our approach here is similar to the one in [6] and [30]. Basically, we decompose the RC algorithm in \mathbb{R}^N into a scalar recursion and a recursion in $\mathbb{R}^{(N-1)}$. We now formally state and prove the result as a theorem.

Theorem 5. *Let $\alpha(t) = a/(t+1)$, $a > 0$, then the RC algorithm in (7) becomes*

$$\mathbf{X}(t+1) = \mathbf{X}(t) + \frac{a}{t+1} [-\boldsymbol{\mu}(\mathbf{X}(t)) + \mathbf{n}(t, \mathbf{X}(t))]. \quad (40)$$

Suppose that the assumptions (A1)-(A5) hold, and that the functions $f(x)$ and $h(x)$ are differentiable with $0 < h'(x) \leq c$, for some $c > 0$. Let the eigenvalue decomposition of \mathbf{L} be given by $\mathbf{L} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T$, where \mathbf{U} is a unitary matrix whose columns are the eigenvectors of \mathbf{L} such that

$$\mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{N}} & \boldsymbol{\Phi} \end{bmatrix}, \boldsymbol{\Phi} \in \mathbb{R}^{N \times (N-1)}, -\boldsymbol{\Sigma} = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{B} \end{bmatrix}, \quad (41)$$

where $\mathbf{B} \in \mathbb{R}^{(N-1) \times (N-1)}$ is a stable diagonal matrix containing the $N-1$ negative eigenvalues of $-\mathbf{L}$ along its diagonal. In addition, let θ_0 be a realization of the random variable θ^ and a is chosen such that $2a\lambda_2(\mathbf{L})g'(0)h'(\theta_0) > 1$ so that the matrix $[ag'(0)h'(\theta_0)\mathbf{B} + \mathbf{I}/2]$, $\theta_0 \in \mathbb{R}$ is stable. Define $[\tilde{n}(t) \ \tilde{\mathbf{n}}(t)^T]^T := N^{-1/2}\mathbf{U}^T\mathbf{n}(t, \mathbf{X}(t))$, $\tilde{\mathbf{n}}(t) \in \mathbb{R}^{(N-1)}$, so that $\tilde{n}(t) = N^{-1}\mathbf{1}^T\mathbf{n}(t, \mathbf{X}(t))$ and $\tilde{\mathbf{n}}(t) = N^{-1/2}\boldsymbol{\Phi}^T\mathbf{n}(t, \mathbf{X}(t))$. Let $\sigma_n^2 := \lim_{t \rightarrow \infty} \text{var}[\tilde{n}(t)]$ and $\mathbf{C} := \lim_{t \rightarrow \infty} \mathbb{E}[\tilde{\mathbf{n}}(t)\tilde{\mathbf{n}}(t)^T]$, $\mathbf{C} \in \mathbb{R}^{(N-1) \times (N-1)}$. Then, as $t \rightarrow \infty$,*

$$\sqrt{t}(\mathbf{X}(t) - \theta_0\mathbf{1}) \sim \mathcal{N}(0, a^2\sigma_n^2\mathbf{1}\mathbf{1}^T + N^{-1}\boldsymbol{\Phi}\mathbf{S}^{\theta_0}\boldsymbol{\Phi}^T), \quad (42)$$

where

$$\mathbf{S}^{\theta_0} = a^2 \int_0^\infty e^{[ag'(0)h'(\theta_0)\mathbf{B} + \mathbf{I}/2]t} \mathbf{C} e^{[ag'(0)h'(\theta_0)\mathbf{B} + \mathbf{I}/2]t} dt. \quad (43)$$

Proof: Define $[\tilde{x}(t) \ \tilde{\mathbf{X}}(t)^T]^T := N^{-1/2}\mathbf{U}^T\mathbf{X}(t)$, $\tilde{\mathbf{X}}(t) \in \mathbb{R}^{(N-1)}$. From Theorem 3, we have $\mathbf{X}(t) \rightarrow \theta^*\mathbf{1}$ a.s. as $t \rightarrow \infty$ which implies that $[\tilde{x}(t) \ \tilde{\mathbf{X}}(t)^T]^T \rightarrow [\theta^* \ \mathbf{0}]^T$ a.s. as $t \rightarrow \infty$, and therefore $\tilde{\mathbf{X}}(t) \rightarrow \mathbf{0}$ a.s. as $t \rightarrow \infty$. For a given θ_0 , the error $[\mathbf{X}(t) - \theta_0\mathbf{1}]$ can be written as the sum of two error components (see also Section VI in [6]) as given below

$$[\mathbf{X}(t) - \theta_0\mathbf{1}] = [\tilde{x}(t) - \theta_0]\mathbf{1} + \frac{1}{\sqrt{N}}\boldsymbol{\Phi}\tilde{\mathbf{X}}(t). \quad (44)$$

Define $\mathbf{e}_1 = [\tilde{x}(t) - \theta_0]\mathbf{1}$ and $\mathbf{e}_2 = N^{-1/2}\Phi\tilde{\mathbf{X}}(t)$ as the first and second terms in (44). By calculating the covariance matrix between \mathbf{e}_1 and \mathbf{e}_2 , it can be proved that they are asymptotically uncorrelated as $t \rightarrow \infty$, and that asymptotically $\sqrt{t}\mathbf{e}_1 \sim \mathcal{N}(0, a^2\sigma_n^2\mathbf{1}\mathbf{1}^T)$ (see Theorem 12 in [6]) where σ_n^2 is the variance of $\tilde{n}(t)$ as $t \rightarrow \infty$ which is calculated to be $\sigma_n^2 = (N^{-2} \sum_{i=1}^N d_i)E_n[f^2(n)]$. To show that $\sqrt{t}\mathbf{e}_2$ is asymptotically normal, it suffices to show that $\sqrt{t}\tilde{\mathbf{X}}(t)$ is asymptotically normal. To this end, we linearize $\boldsymbol{\mu}(\mathbf{x})$ in (40) around $\mathbf{x} = \theta_0\mathbf{1}$ using Taylor's series expansion,

$$\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\mu}(\theta_0\mathbf{1}) + \left. \frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\theta_0\mathbf{1}} (\mathbf{x} - \theta_0\mathbf{1}) + o(\|\mathbf{x} - \theta_0\mathbf{1}\|), \quad (45)$$

$$= g'(0)h'(\theta_0)\mathbf{L}\mathbf{x} + o(\|\mathbf{x} - \theta_0\mathbf{1}\|), \quad (46)$$

where the Jacobian matrix of $\boldsymbol{\mu}(\mathbf{x})$ has i, j element given by $\left[\frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \mathbf{x}} \right]_{i,j} = \frac{\partial \mu_i(\mathbf{x})}{\partial x_j}$.

Using (46) in (40) we get

$$\mathbf{X}(t+1) = \mathbf{X}(t) + \frac{a}{t+1} \left[g'(0)h'(\theta_0)(-\mathbf{L}\mathbf{X}(t)) + o(\|\mathbf{X}(t) - \theta_0\mathbf{1}\|) + \mathbf{n}(t, \mathbf{X}(t)) \right], \text{ as } t \rightarrow \infty. \quad (47)$$

Pre-multiplying (47) on both sides by $N^{-1/2}\mathbf{U}^T$ and using (41) we get the following recursions

$$\tilde{x}(t+1) = \tilde{x}(t) + \frac{a}{t+1}\tilde{n}(t), \quad (48)$$

$$\tilde{\mathbf{X}}(t+1) = \tilde{\mathbf{X}}(t) + \frac{a}{t+1} \left[g'(0)h'(\theta_0)\mathbf{B}\tilde{\mathbf{X}}(t) + o(\|\mathbf{X}(t) - \theta_0\mathbf{1}\|) + \tilde{\mathbf{n}}(t) \right], \text{ as } t \rightarrow \infty. \quad (49)$$

In [33], asymptotic normality of a recursion similar to (49) has been proved under certain conditions. With the assumption that $[ag'(0)h'(\theta_0)\mathbf{B} + \mathbf{I}/2]$ is a stable matrix for $\theta_0 \in \mathbb{R}$, it can be verified that all the conditions of Theorem 6.6.1 in [33, p. 147] are satisfied for the process $\tilde{\mathbf{X}}(t)$ in (49). Therefore, for a given θ_0 , the process $\sqrt{t}\tilde{\mathbf{X}}(t)$ is asymptotically normal with zero mean and covariance matrix given by (43). Since $\sqrt{t}\mathbf{e}_1 \sim \mathcal{N}(0, a^2\sigma_n^2\mathbf{1}\mathbf{1}^T)$ and using (43) together with the fact that \mathbf{e}_1 and \mathbf{e}_2 are asymptotically independent as $t \rightarrow \infty$, we get (42) which completes the proof. \blacksquare

Equation (42) indicates how fast the process $\mathbf{X}(t)$ will converge to $\theta_0\mathbf{1}$ for a given θ_0 . The convergence speed clearly depends on $g'(0)$ and $h'(\theta_0)$ which captures the effect of receiver and transmit non-linearities respectively.

Let the asymptotic covariance in (42) be denoted by \mathbf{C}_{RC} . Since $\tilde{\mathbf{n}}(t)$ are asymptotically i.i.d. across space and time, \mathbf{C} in (43) becomes $\mathbf{C} = N\sigma_n^2\mathbf{I}$ with $\sigma_n^2 = (N^{-2} \sum_{i=1}^N d_i)E_n[f^2(n)]$ and

thus we have $\mathbf{C}_{\text{RC}} = a^2 \sigma_n^2 \mathbf{1}\mathbf{1}^T + N^{-1} \Phi \mathbf{S}^{\theta_0} \Phi^T$ where \mathbf{S}^{θ_0} is a diagonal matrix whose diagonal elements are given by $\mathbf{S}_{i,i}^{\theta_0} = a^2 \sigma_n^2 / [2ag'(0)h'(\theta_0)\lambda_{i+1}(\mathbf{L}) - 1]$. A reasonable quantitative measure of largeness [35] of the asymptotic covariance matrix is $\|\mathbf{C}_{\text{RC}}\|$ which is the maximum eigenvalue of the symmetric matrix \mathbf{C}_{RC} .

Further, $\|\mathbf{C}_{\text{RC}}\|$ can be minimized with respect to the parameter a and this can be formulated as the following optimization problem,

$$\min_{\{a | 2ag'(0)h'(\theta_0)\lambda_2(\mathbf{L}) > 1\}} \max_{\{\mathbf{x} | \mathbf{x} \in \mathbb{R}^N, \|\mathbf{x}\|^2 \leq 1\}} \mathbf{x}^T \mathbf{C}_{\text{RC}} \mathbf{x}, \quad (50)$$

which can be solved analytically. The value of a that optimizes (50) is found to be $a_{\text{RC}}^* = (N+1)/[2N\lambda_2(\mathbf{L})g'(0)h'(\theta_0)]$ and the corresponding optimal value of the $\|\mathbf{C}_{\text{RC}}\|$ is given by

$$\|\mathbf{C}_{\text{RC}}^*\| = \left(N^{-2} \sum_{i=1}^N d_i \right) \left(\frac{N+1}{2N} \right)^2 \left(\frac{\mathbb{E}_n[f^2(n)]}{(\mathbb{E}_n[f'(n)])^2} \right) \left(\frac{1}{\lambda_2^2(\mathbf{L})} \right) \left(\frac{1}{h'(\theta_0)} \right)^2. \quad (51)$$

The best speed of convergence characterized by the asymptotic covariance depends on the point of convergence through $h'(\theta_0)$. To select the optimal a that would result in the best speed of convergence for a given $f(x)$ and $h(x)$, knowledge of θ_0 is required. Since θ_0 unknown apriori, the performance characterized in (51) could serve as the benchmark for a given $h(x)$. In practice it may be possible for sensors to adapt the value of a as they converge towards the limiting value θ_0 to speed up the convergence, and approach this benchmark. An optimized value for a , also provides a simpler final expression for the asymptotic covariance in terms of its dependence on the receive nonlinearity $f(x)$ and transmit nonlinearity $h(x)$.

The size of the asymptotic covariance matrix in (51) is inversely proportional to the square of the smallest non-zero eigenvalue $\lambda_2(\mathbf{L})$ which quantifies how densely a graph is connected. Even though the asymptotic covariance \mathbf{C}_{RC} has been derived in the literature [6], its optimization has not been considered. The optimization considered in (50) enables us to infer some interesting conclusions. In Table I, we have summarized the behavior of $\|\mathbf{C}_{\text{RC}}^*\|$ for several graphs for large N [36]–[39]. For the fully connected graph, $\|\mathbf{C}_{\text{RC}}^*\|$ goes to zero faster than the star graph and thus the former will converge faster than the latter. For the ring and line graphs, with large N , the convergence will become slower since $\|\mathbf{C}_{\text{RC}}^*\|$ increases with N . For other graphs in Table I, the convergence speed is better compared to the line and ring graphs since $\|\mathbf{C}_{\text{RC}}^*\|$ decreases with N for those graphs. It is also interesting to note that the minimization of (51) with respect to the transmit and receive nonlinearities can be done separately and thus asymptotic

Type of Graph	$\lambda_2(\mathbf{L})$	Behavior of $\ \mathbf{C}_{\text{RC}}^*\ $
Fully Connected	N	$O(N^{-2})$
Star	1	$O(N^{-1})$
Ring	$4 \sin^2\left(\frac{\pi}{N}\right)$	$O(N^3)$
Line	$4 \sin^2\left(\frac{\pi}{2N}\right)$	$O(N^3)$
Tree (excluding star graphs)	≤ 0.3819	$O(N^{-1})$
Cubic Graph	2	$O(N^{-1})$
Planar	≤ 4	$O(N^{-1})$
Bipartite complete graph with p and q vertices	$\min(p, q)$	$O(N^{-1})$
k -regular (includes Ramanujan graphs)	$\leq k - 2\sqrt{k-1}$	$O(N^{-1})$
k -regular Lattice	$(k+1) - \frac{\sin((k+1)\frac{\pi}{N})}{\sin(\frac{\pi}{N})}$	$O\left(N^{-1} \left((k+1) - \frac{\sin((k+1)\frac{\pi}{N})}{\sin(\frac{\pi}{N})}\right)^{-2}\right)$

TABLE I
BEHAVIOR OF $\|\mathbf{C}_{\text{RC}}^*\|$ FOR SOME COMMON GRAPHS

covariance is an easier and helpful metric in optimizing the performance. The nonlinear receiver function $f(x)$ for which the ratio $E_n[f^2(n)]/(E_n[f'(n)])^2$ is smaller will be better in terms of speed of convergence. For example, if n is Laplacian distributed with variance of 2 and if $f(x) = x$, then $E_n[f^2(n)]/(E_n[f'(n)])^2 = 2$ whereas if we choose $f(x) = \tanh(x)$, we have $E_n[f^2(n)]/(E_n[f'(n)])^2 = 1.317$ indicating $\tanh(x)$ will perform better than the linear case. This is due to the fact that Laplacian is a heavy tailed distribution and therefore a bounded function such as $\tanh(x)$ curtails the effect of outliers which does not happen when $f(x)$ is linear. Equation (51) also indicates when $h(x)$ is fixed, scaling $f(x)$ does not change the speed of convergence. We will illustrate these findings using simulations in Section V.

When $f(x)$ is a bounded function, from equation (8) in [40] we have

$$\frac{E_n[f^2(n)]}{(E_n[f'(n)])^2} \geq \frac{1}{J}, \quad (52)$$

where J is the Fisher information of n with respect to a location parameter [41, (8)] and thus we see an interesting relationship between the maximum eigenvalue of the asymptotic covariance and the Fisher information. For any $h(x)$, the best choice of $f(x)$ is the one that achieves equality in (52). For instance, when n is Gaussian, $f(x) = x$ achieves equality in (52) in which case we have $\text{var}[f(n)]$ equals the inverse of Fisher information. In addition, when n has finite moments, our RC algorithm in (4) subsumes the non-linear consensus algorithm discussed in [30] with $f(x) = x$, and we get the same result as in (51) except $E_n[f^2(n)]$ is replaced by the noise

variance σ_v^2 defined in [30]. Further, our model subsumes the linear case studied in [6] with $f(x) = x$ and $h(x) = x$.

V. SIMULATIONS

In this section, we corroborate our analytical findings through various simulations. In all the simulations presented, the initial samples $x_i(0) \in \mathbb{R}, i = 1, 2, \dots, N$, were generated randomly using Gaussian distribution with a standard deviation equal to 10. The desired global average value is indicated in each of the simulations. We focus here on bounded functions for both the transmit and receiver non-linearities to study their performance.

A. Performance of RC algorithm with Channel Noise

First, we highlight that the linear consensus algorithms in [6], [8]–[16] fail to achieve consensus when the channel noise does not have finite variance. An example plot is shown in Figure 1 for the case when the channel noise is Cauchy distributed with the scale parameter $\gamma = 1$. Clearly, the sensors do not reach consensus. Whereas the proposed RC algorithm works when we choose $f(x)$ as a nonlinear function as shown next.

Figures 2 - 7 illustrate the performance of RC algorithm in the presence of communication noise. As explained in the assumption **(A5)** in Section IV-A, we chose the decreasing step sequence to be $\alpha(t) = 1/(t+1), t \geq 0$, in all simulations. Here we assumed that $\rho = \max_x h^2(x)$ is the maximum power available at each sensor to transmit its state value. The receiver nonlinear function $f(x)$ is indicated in each case. Figure 2 shows that the nodes employing the RC algorithm reach consensus for a small network with $N = 10$ in about 100 iterations and Figure 3 shows convergence for a large network with $N = 75$ in about 40 iterations.

In Figures 4, 5 and 6 we show the convergence speed performance of the proposed RC algorithm by plotting the maximum eigenvalue of the covariance matrix of the vector process $\sqrt{t}(\mathbf{X}(t) - \theta_0 \mathbf{1})$ versus iterations t . These plots indicate how fast the process $\mathbf{X}(t)$ converges towards the limiting value $\theta_0 \mathbf{1}$.

The speed of convergence for two graphs with different algebraic connectivity is illustrated in Figure 4. We see that the graph with smaller connectivity (smaller $\lambda_2(\mathbf{L})$) converges slower than the one with large connectivity as dictated by (51). In Theorem 5, we also saw that scaling $f(x)$ does not change the asymptotic convergence speed. This is shown in Figure 5 where we see

that when the iterations are large ($t > 140$), the speed of convergence of all the three functions are nearly the same. We depict the robustness of the RC algorithm for various channel noise distributions in Figure 6. We observe that the performance is nearly the same for Gaussian and Laplacian distributions, whereas there is a significant gap between Cauchy and alpha-stable distributions considered in this simulation. The latter effect is due to the fact that, for a given $f(\cdot)$, the ratio $E_n[f^2(n)]/(E_n[f'(n)])^2$ is significantly different for those two cases which justifies the performance gap. Finally, we illustrate the difference between the variance of θ^* and the asymptotic variance in Figure 7. Here we consider the evolution of the state value $x_1(t)$ of the first node for several consensus runs for the same initial conditions. Recall that in every consensus run the state value $x_1(t)$ converges to an instance of the limiting random variable θ^* and the variation among these several realizations is characterized by the variance of θ^* . In contrast, how fast the state value $x_1(t)$ converges to the limiting value θ_0 is characterized by the asymptotic variance of $\sqrt{t}[x_1(t) - \theta_0]$ as $t \rightarrow \infty$.

VI. CONCLUSIONS

A distributed average consensus algorithm that converges in the presence of impulsive noise is considered. Every sensor also maps its state value through a bounded function before transmission to constrain the transmit power. It is shown that non-linearity at the receiver nodes makes the algorithm robust to a wide range of channel noise distributions including heavy-tailed channel noise. The proposed algorithm relaxes the requirement of finite moments on the communication noise and thus it is proved to be not only more general than the existing consensus algorithms but is practically viable for WSNs deployed in adverse conditions. It is proved using the theory of Markov processes that the sensors reach consensus asymptotically on a finite random variable whose expectation contains the desired sample average of the initial sensor measurements, and whose mean-squared error is bounded. The asymptotic convergence speed of the proposed algorithm is characterized by deriving the asymptotic covariance matrix using results from stochastic approximation theory. It is shown that the norm of the asymptotic covariance matrix is limited by the Fisher information of the noise distribution with respect to a location parameter.

REFERENCES

- [1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *Communications Magazine*, *IEEE*, vol. 40, no. 8, pp. 102 – 114, Aug 2002.

- [2] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 5, Dec. 2003, pp. 4997 – 5002.
- [3] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Analysis and optimization of randomized gossip algorithms," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 5, Dec. 2004, pp. 5310 – 5315.
- [4] R. Saber and R. Murray, "Consensus protocols for networks of dynamic agents," in *American Control Conference, 2003. Proceedings of the 2003*, vol. 2, Apr. 2003, pp. 951 – 956.
- [5] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215 –233, Jan. 2007.
- [6] M. Huang and J. Manton, "Stochastic consensus seeking with measurement noise: Convergence and asymptotic normality," in *American Control Conference, 2008*, June 2008, pp. 1337 –1342.
- [7] B. Oreshkin, T. Aysal, and M. Coates, "Distributed average consensus with increased convergence rate," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, Mar. 2008, pp. 2285 –2288.
- [8] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 355 –369, Jan. 2009.
- [9] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *J. Parallel Distrib. Comput.*, vol. 67, pp. 33–46, Jan. 2007. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1222667.1222952>
- [10] B. Touri and A. Nedic, "Distributed consensus over network with noisy links," in *Information Fusion, 2009. FUSION '09. 12th International Conference on*, July 2009, pp. 146 –154.
- [11] M. Huang and J. Manton, "Stochastic approximation for consensus seeking: Mean square and almost sure convergence," in *Decision and Control, 2007 46th IEEE Conference on*, Dec. 2007, pp. 306 –311.
- [12] L. Pescosolido, S. Barbarossa, and G. Scutari, "Average consensus algorithms robust against channel noise," in *Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 9th Workshop on*, July 2008, pp. 261 –265.
- [13] S. Barbarossa, T. Battisti, L. Pescosolido, S. Sardellitti, and G. Scutari, "Distributed processing algorithms for wireless sensor networks having fast convergence and robustness against coupling noise," in *Spread Spectrum Techniques and Applications, 2008. ISSSTA '08. IEEE 10th International Symposium on*, Aug. 2008, pp. 1 – 6.
- [14] T. Aysal and K. Barner, "Convergence of consensus models with stochastic disturbances," *Information Theory, IEEE Transactions on*, vol. 56, no. 8, pp. 4101 –4113, Aug. 2010.
- [15] K. Srivastava and A. Nedic, "Distributed asynchronous constrained stochastic optimization," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 4, pp. 772–790, 2011.
- [16] S. Kar and J. Moura, "Distributed average consensus in sensor networks with random link failures and communication channel noise," in *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, Nov. 2007, pp. 676 –680.
- [17] E. Sousa, "Performance of a spread spectrum packet radio network link in a Poisson field of interferers," *IEEE Trans. Inf. Theory*, vol. 38, no. 6, pp. 1743–1754, Nov. 1992.
- [18] J. Ilow and D. Hatzinakos, "Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers," *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1601–1611, Jun. 1998.
- [19] X. Yang and A. Petropulu, "Co-channel interference modeling and analysis in a Poisson field of interferers in wireless communications," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 64–76, Jan. 2003.
- [20] B. Hughes, "Alpha-stable models of multiuser interference," in *Proc. IEEE ISIT'00*, Jun. 2000, p. 383.

- [21] M. Haenggi, J. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, “Stochastic geometry and random graphs for the analysis and design of wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1029–1046, Sep. 2009.
- [22] M. Win, P. Pinto, and L. Shepp, “A mathematical theory of network interference and its applications,” *Proc. IEEE*, vol. 97, no. 2, pp. 205–230, Feb. 2009.
- [23] J. Lee and C. Tepedelenlioglu, “Space-time coding over fading channels with stable noise,” *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 7, pp. 3169–3177, 2011.
- [24] A. Rajan and C. Tepedelenlioglu, “Diversity combining over rayleigh fading channels with symmetric alpha-stable noise,” *Wireless Communications, IEEE Transactions on*, vol. 9, no. 9, pp. 2968–2976, 2010.
- [25] U. Khan, S. Kar, and J. Moura, “Distributed average consensus: Beyond the realm of linearity,” in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, nov. 2009, pp. 1337 – 1342.
- [26] U. Munz, A. Papachristodoulou, and F. Allgower, “Nonlinear multi-agent system consensus with time-varying delays,” *Proceedings of the 17th World Congress The International Federation of Automatic Control*, pp. 6 – 11, July 2008.
- [27] Q. Hui and W. M. Haddad, “Distributed nonlinear control algorithms for network consensus,” *Automatica*, vol. 44, no. 9, pp. 2375 – 2381, 2008.
- [28] W. Yu, G. Chen, and M. Cao, “Consensus in directed networks of agents with nonlinear dynamics,” *Automatic Control, IEEE Transactions on*, vol. 56, no. 6, pp. 1436–1441, 2011.
- [29] A. Ajorlou, A. Momeni, and A. G. Aghdam, “Sufficient conditions for the convergence of a class of nonlinear distributed consensus algorithms,” *Automatica*, vol. 47, no. 3, pp. 625 – 629, 2011.
- [30] S. Dasarathan, C. Tepedelenliolu, M. Banavar, and A. Spanias, “Non-linear distributed average consensus using bounded transmissions,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 6000–6009, Dec 2013.
- [31] F. Chung, *Spectral Graph Theory*, ser. Regional Conference Series in Mathematics. Conference Board of the Mathematical Sciences, 1997, no. 92.
- [32] R. Olfati-Saber and R. Murray, “Consensus problems in networks of agents with switching topology and time-delays,” *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1520 – 1533, Sept. 2004.
- [33] M. Nevelson and R. Khasminskii, *Stochastic Approximation and Recursive Estimation*. Amer. Math. Soc., 1973.
- [34] D. Williams, *Probability with Martingales*, ser. Cambridge Mathematical Textbooks. Cambridge University Press, 1991. [Online]. Available: <http://books.google.com/books?id=e9saZ0YSi-AC>
- [35] B. Polyak and Y. Tsypkin, “Optimal pseudogradient adaptation procedures,” *Automat, Remote Control*, vol. 41, pp. 1101–1110, 1981.
- [36] N. M. M. de Abreu, “Old and new results on algebraic connectivity of graphs,” *Linear Algebra and its Applications*, vol. 423, no. 1, pp. 53 – 73, 2007, special Issue devoted to papers presented at the Aveiro Workshop on Graph Spectra Aveiro Workshop on Graph Spectra. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0024379506003971>
- [37] X.-D. Zhang, “The laplacian eigenvalues of graphs: a survey,” 2011. [Online]. Available: <http://arxiv.org/abs/1111.2897>
- [38] R. Olfati-Saber, “Algebraic connectivity ratio of ramanujan graphs,” in *American Control Conference, 2007. ACC '07*, July 2007, pp. 4619–4624.
- [39] D. A. Spielman, “The laplacian,” 2012. [Online]. Available: <http://www.cs.yale.edu/homes/spielman/561/lect02-12.pdf>
- [40] Z. Zhang, “Inequalities for characteristic functions involving Fischer information,” *Comptes Rendus Mathematique*, vol. 344, no. 5, pp. 327–330, March 2007.
- [41] R. Zamir, “A proof of the Fisher information inequality via a data processing argument,” *Information Theory, IEEE Transactions on*, vol. 44, no. 3, pp. 1246–1250, May 1998.

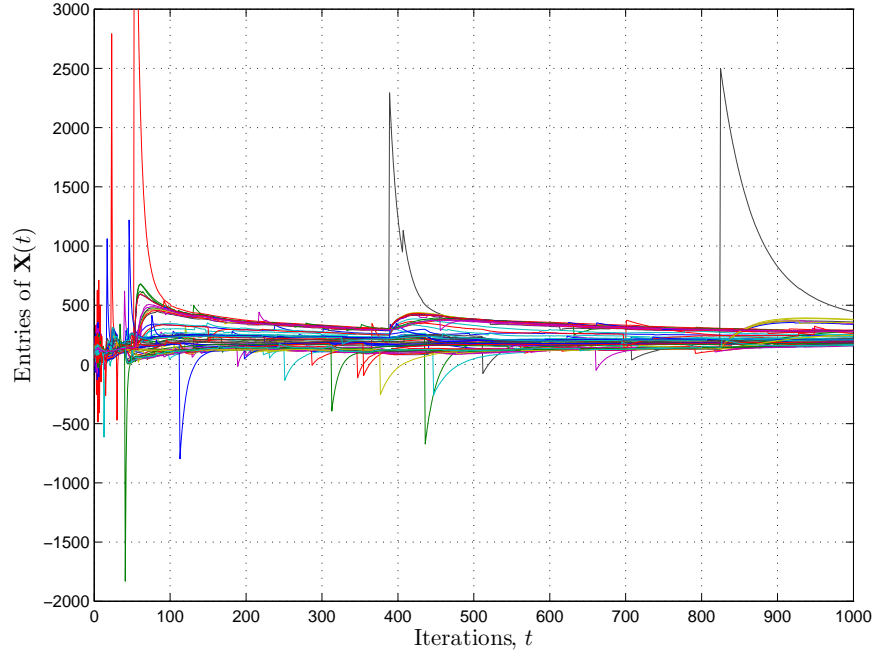


Fig. 1. Linear consensus fails with impulsive channel noise, Entries of $\mathbf{X}(t)$ versus Iterations t : Cauchy noise, $h(x) = x$, $f(x) = x$, $N = 75$, $\bar{x} = 134.31$, $\gamma = 1$.

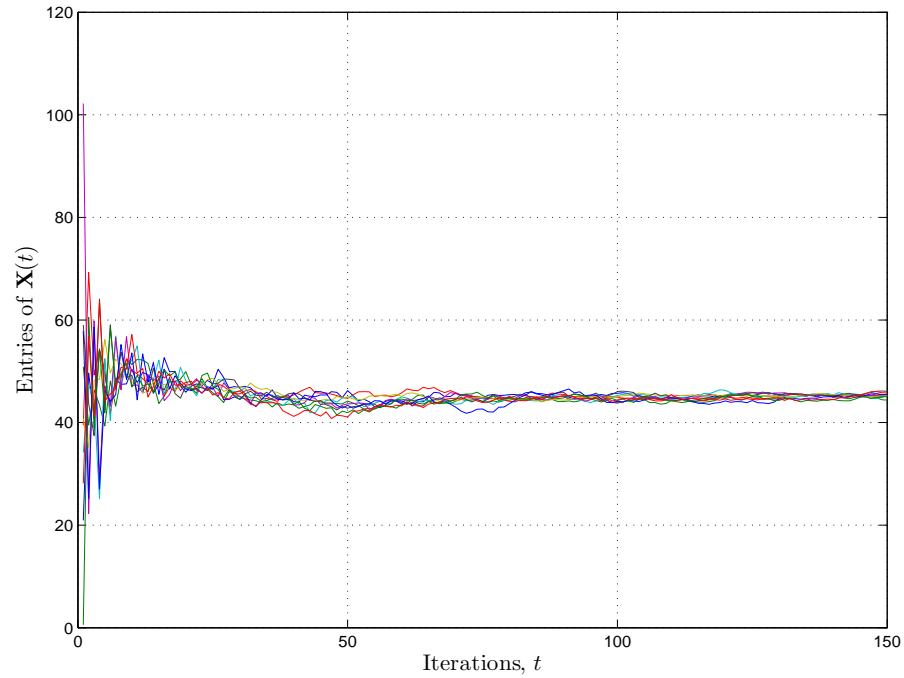


Fig. 2. Convergence of RC algorithm for a small Graph, Entries of $\mathbf{X}(t)$ versus Iterations t : Cauchy noise, $h(x) = \sqrt{\rho} \frac{2}{\pi} \tan^{-1}(\frac{\pi}{2} 0.01x)$, $f(x) = \tanh(5x)$, $N = 10$, $\bar{x} = 43.96$, $\rho = 15$ dB, $\gamma = 0.1$.

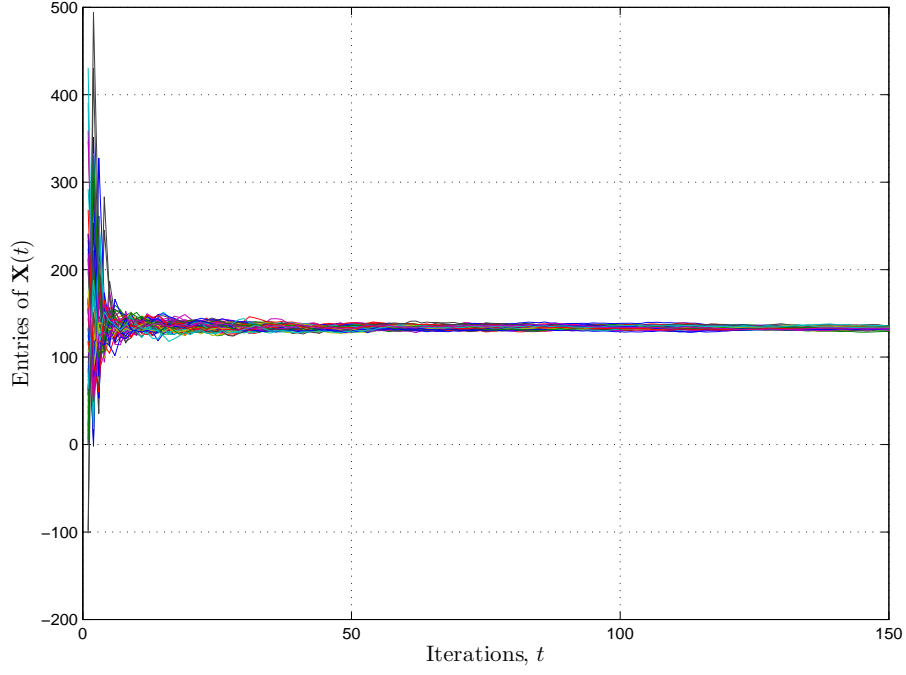


Fig. 3. Convergence of RC algorithm for a large Graph, Entries of $\mathbf{X}(t)$ versus Iterations t : Cauchy noise, $h(x) = \sqrt{\rho} \frac{2}{\pi} \tan^{-1}(\frac{\pi}{2} 0.01x)$, $f(x) = \tanh(5x)$, $N = 75$, $\bar{x} = 134.31$, $\rho = 5$ dB, $\gamma = 0.1$.

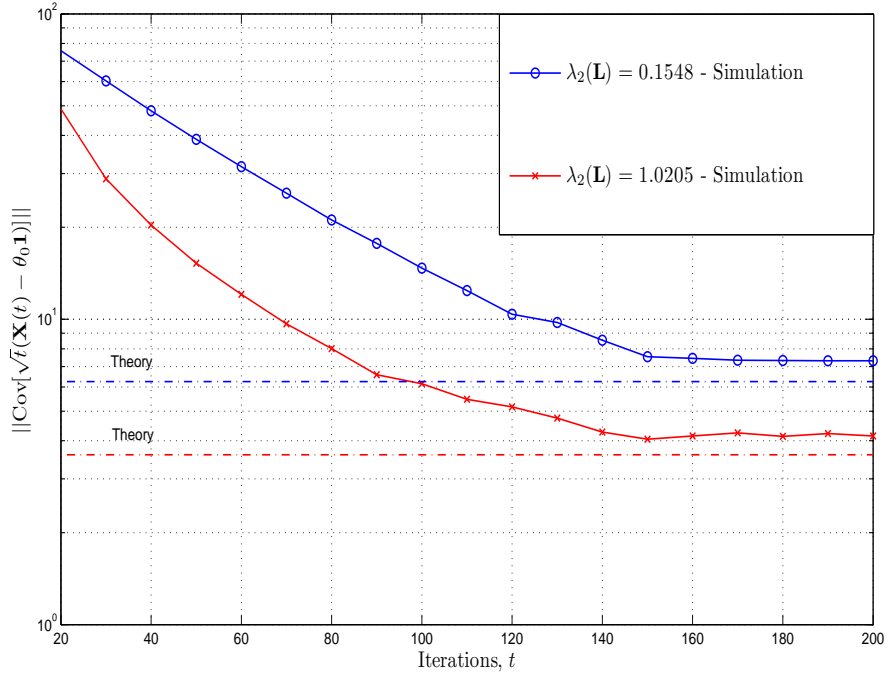


Fig. 4. Difference in speed of convergence: sparsely versus densely connected Graphs, $\|\text{Cov}[\sqrt{t}(\mathbf{X}(t) - \theta_0 \mathbf{1})]\|$ versus Iterations t : Cauchy noise, $h(x) = x$, $f(x) = \frac{1.5x}{1+|1.5x|}$, $N = 75$, $\theta_0 = 85.49$, $\bar{x} = 84.31$, $\gamma = 0.413$.

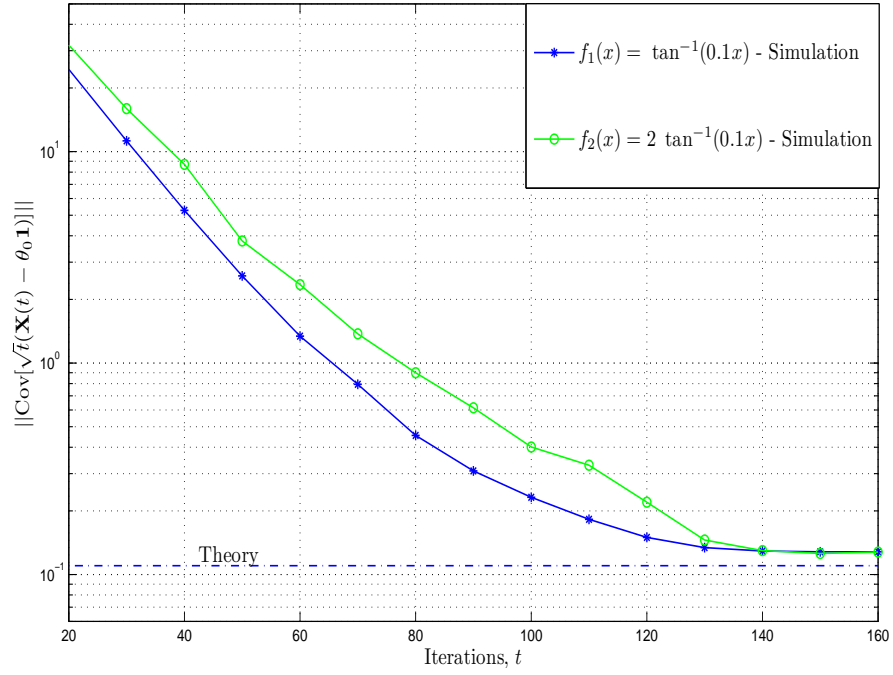


Fig. 5. Scaling $f(x)$ does not change speed of convergence, $\|\text{Cov}[\sqrt{t}(\mathbf{X}(t) - \theta_0 \mathbf{1})]\|$ versus Iterations t : $h(x) = x$, $N = 10$, $\theta_0 = 32.63$, $\bar{x} = 34.31$, $\gamma = 0.413$.

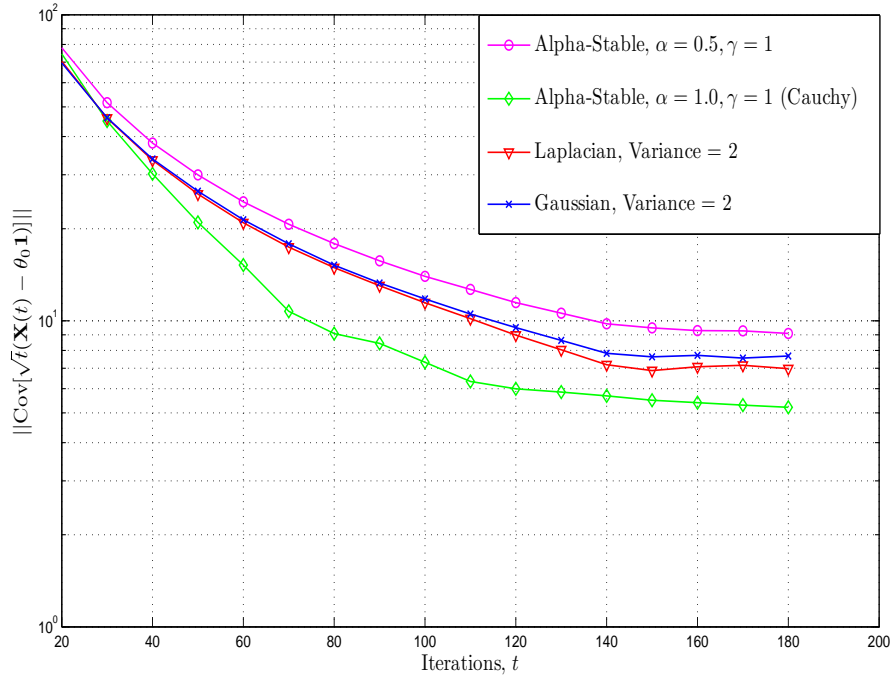


Fig. 6. Robustness to various noise distributions, $\|\text{Cov}[\sqrt{t}(\mathbf{X}(t) - \theta_0 \mathbf{1})]\|$ versus Iterations t : $h(x) = x$, $f(x) = \tanh(2x)$, $N = 75$, $\theta_0 = 120.36$, $\bar{x} = 124.31$

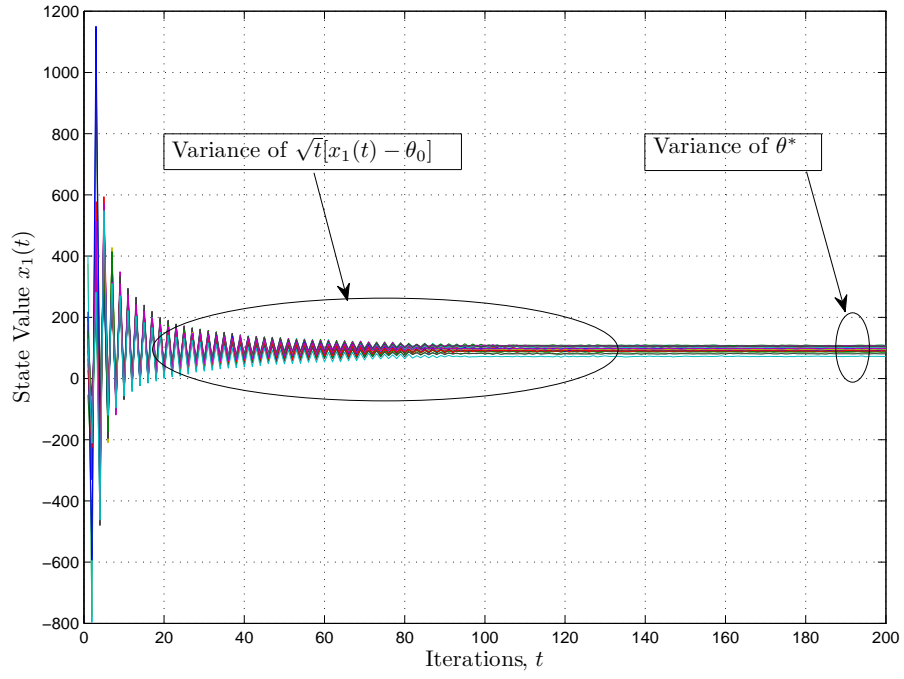


Fig. 7. Difference between Variance of θ^* and Asymptotic Variance, Entries of $\mathbf{X}(t)$ versus Iterations t : $h(x) = x$, $f(x) = 3 \tan^{-1}(0.05x)$, $N = 75$, $\bar{x} = 94.31$, $\gamma = 0.413$